DOCUMENT RESUME

ED 361 346 TM 020 392

AUTHOR Powell, Douglas A.
TITLE An Extension of the Covariate Randomized Response Model to the Case of a Nonmonotonic Covariate.
PUB DATE 93
NOTE 9p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Administrators; Business; Equations (Mathematics); *Error of Measurement; *Estimation (Mathematics); *Mathematical Models
IDENTIFIERS *Covariation; Nonmonotonic Covariates; *Randomized Response Technique

ABSTRACT

The use of a covariate for randomized response (RRT) research has been shown to reduce standard errors of sensitive trait proportion estimates. At the same time, the model has been shown to be subject to serious misspecification when the relationship between the covariate and the sensitive trait is non-monotonic. The RRT covariate model is adapted to the non-monotonic case, and the efficacy of this approach is demonstrated through the analysis of a set of data from a business school study of bluffing behavior by business executives. Approximately 500 executives from a broad range of industries participated in a study of their tendencies to bluff when announcing a price increase. While the study does not develop a general non-monotonic covariate RRT model, it shows the potential utility of alternatives to the standard covariate approach and the reasonableness of attempting a non-monotonic model through the demonstrated capacity of the merged logistic function to estimate the sensitive trait proportions more precisely. Three tables present analysis results. (Contains 5 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 361 346

AN EXTENSION OF THE COVARIATE RANDOMIZED RESPONSE MODEL TO THE CASE OF A NONMONOTONIC COVARIATE.

DOUGLAS A. POWELL
UNIVERSITY OF MARYLAND

The use of a covariate for randomized response (RRT) research has been shown to reduce standard errors of sensitive trait proportion estimates. At the same time, the model has been shown to be subject to serious misspecification when the relationship between the covariate and the sensitive trait is nonmonotonic. This paper adapts the RRT covariate model to the nonmonotonic case. The efficacy of this approach is demonstrated through the analysis of a set of data from a business school study of bluffing behavior by business executives.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN **OTHER THAN PAPER COPY** HAS BEEN GRANTED BY

Douglas A. Powell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

TM020392

BEST COPY AVAILABLE

1. INTRODUCTION

The randomized response technique (RRT) was developed by Warner (1965) to obtain answers to sensitive questions in survey research. The logic of the technique involves the presentation of pairs of questions to the respondent. The respondent randomly chooses which member of the pair to answer and thus the particular question responded to cannot be known by the researcher. Given frequencies of "yes" and "no" responses and the probability of selecting the sensitive question (this is fixed by the choice of the randomizing device, e.g., a spinner) the population proportion for the sensitive trait and its standard error can be estimated from the sample responses.

The RRT procedure has been shown to reduce response bias in comparison with anonymous questionaires (Sudman and Bradburn, 1974). On the other hand, the standard errors of the sensitive trait proportions are larger under RRT due to the randomization of responses. Scheers and Dayton (1988) demonstrated that the use of a covariate that correlates with the sensitive trait can lead to the reduction in standard errors for estimates of population proportions for sensitive traits. However, they found that the covariate model may be seriously misspecified when the relationship between the covariate and the sensitive trait is nonmonotonic. This paper describes an extension of the covariate model to the case of a covariate that is nonmonotonically related to the sensitive trait.

After reviewing the history of these models, this paper will present the underlying mathematics of the nonmonotonic model. Additionally its efficacy will be demonstrated relative to the monotonic covariate model using contrived data. The two models will also be compared for data obtained from a business school study of bluffing behavior by business executives.

2. HISTORICAL REVIEW

The Warner (1965) model requires the presentation of a sensitive question in an affirmative and a negative form. The respondent is instructed to answer one member of the pair based on the outcome of the use of a randomizing device such as a spinner. Other models have been developed that employ the "unrelated" question format (Greenberg et al., 1969), i.e., the sensitive question is paired with a question that is unrelated to the sensitive question. Consider the following:

Warner Model:	Unrelated Question Model:
A. I have used illegal drugs.	A. I have used illegal drugs.
B. I have never used illegal drugs.	B. The first digit of my phone number is less than 5.

While the unrelated question procedure is widely used in RRT based surveys, I will describe further developments in terms of the Warner model only. The techniques described herein are readily adaptable to the unrelated question model.

The investigator fixes the probability of responding to the sensitive question by the choice of the randomizing device and collects the frequencies of "yes" responses to each item for each respondent. For the standard Warner technique, the probability of a "yes" response is modeled as such:

$$P(Y) = p\pi_a + (1 - p)(1 - \pi_a). \qquad (2.1)$$

In (2.1) above the probability of selecting the sensitive question is given by p. The proportion of the population having the sensitive trait is given by π_a. In the covariate extension of the Warner model (Scheers and Dayton, 1988) the researcher must collect a covariate variable, X, for each survey respondent. The presence of the covariate requires the modification of (2.1) to reflect the fact that the

conditional probability of a "yes" response is being modeled. The covariate model for the proportion of persons with the sensitive trait conditional on the covariate level X is given by:

$$\pi_{a|X} = g(X) \tag{2.2}$$

where g(X) is a logistic function of X. The logistic function has two parameters, α and β, and is given by:

$$g(X) = 1/[1 + \exp(-\alpha - \beta X)]. \tag{2.3}$$

The conditional probability of a "yes" response given the covariate level X, is then given by:

$$P(Y|X) = pg(X) + (1 - p)[1 - g(X)]. \tag{2.4}$$

The parameter estimates for the logistic model, i.e., the α and the β parameters, are obtained via maximum likelihood estimation. The resulting parameter estimates can be used to estimate the proportion of "yes" responses at each covariate level and the proportion of individuals with the sensitive trait at each covariate level. Scheers and Dayton (1988) developed a program to produce these estimates as well as the standard errors of the trait proportions.

The covariate model was shown to give lower standard errors for sensitive trait proportions than the Warner model for cases involving a monotonic covariate. The potential for model misspecification clearly exists if the sensitive trait follows a nonmonotonic relationship with the covariate. (Scheers and Dayton, 1988).

3. THE NONMONOTONIC COVARIATE MODEL

To permit the usage of the covariate model for situations involving an underlying nonmonotonic relationship between the sensitive trait proportion and the covariate, the standard logistic function must be modified. The logistic function has the characteristic of mathematical simplicity and conforms to the requirement of being limited to the (0,1) interval (a requirement for any function that models probabilities).

The proposed modification involves the designation of a "merge" point, μ, such that slope of the logistic curve (β) changes sign at μ. An additional restriction is that the slope of the "merged" logistic function has the same absolute value on either side of μ. The modified covariate model has the following functional form:

$$g(X) = 1/\{1 + \exp[-\alpha - j\beta(X - \mu)]\}. \tag{3.1}$$

where j represents a sign term, i.e., j equals -1 if X is less than μ and is equal to 1 otherwise.

The corresponding expression for the probability of a "yes" response for a given level of the covariate, X, is given by:

$$P(Y|X) = pg(X) + (1 - p)(1 - g(X)). \tag{3.2}$$

where the randomizing proportion is given by p as before and is controlled by the researcher. A typical value for p is .7.

If the covariate assumes c levels in the sample, an expression for the likelihood of the data is given by:

$$L = \prod_{i=1}^{c} P(Y|X_i)^{y_i}[1 - P(Y|X_i)]^{n_i} \tag{3.3}$$

where y_i and n_i represent the frequencies of yes and no responses respectively at the i^{th} level of the covariate. The estimation of parameters is done by maximum likelihood estimation. The use of an iterative procedure such as the Newton-Raphson algorithm is typically required, because the first and second partial derivatives of the likelihood function only rarely give normal equations that are amenable to an analytic solution. The present application utilizes the Fisher scoring method (Rao, 1965) to obtain parameter estimates.

The merge point, μ, is determined by a guided search technique. In this technique the abscissa of the covariate axis is searched for a merge point that maximizes the likelihood function associated with a set of parameter estimates. In other words, a likelihood function is maximized for each potential merge point and the "maximum of maxima" is selected as the optimum merge point. The parameter estimates obtained for the optimum merge point are inserted into (3.1) to obtain an estimate of the proportion of persons with the sensitive trait at each covariate level. The estimation process requires the election by the researcher of an increment for the guided search. For this application an increment of .05 of a unit is employed.

The overall goodness of fit for the model is assessed by a chi-square goodness of fit test. Either a Pearson or a likelihood ratio chi-square test can be employed (Rao, 1965). The author has developed a program, written in the GAUSS programming language, to obtain parameter estimates, their associated standard errors, the chi-sqaure tests of fit, the estimates of sensitive trait proportions at each covariate level and the standard errors of these estimated proportions. The overall weighted average of the estimates of sensitive trait proportions (across all levels of the covariate) is also computed along with the standard error of this weighted average. The weighted average is computed based on weights equal to the number of observations at each covariate level divided by the total number of observations in the sample.

4. USE OF MODEL WITH CONTRIVED DATA

Scheers and Dayton (1988) found that their monotonic model was in jeopardy of misspecification when the sensitive trait varied with the covariate in other than a monotonic fashion. They provided the following data set as an example of such a misspecification:

Cov. Level	# Yes	# No
1	4	9
2	20	28
3	39	39
4	20	28
5	4	9

The same data was analyzed with the nonmonotonic covariate technique. The probability of selecting the sensitive item was set at .70. Whereas the standard covariate model produced a parameter estimate for the slope, $\hat{\beta} = .000$, indicating model misspecification, the nonmonotonic model yields $\hat{\beta} = -1.117$ with a standard error of .342. The ratio of the slope term to its standard error gives a z-statistic (approximate) of 3.27, indicating that the "merged" logistic covariate model permits detection of an underlying relationship between the covariate and the sensitive trait proportion. Additionally, the likelihood-ratio chi-square statisitc was .198 with 2 df, indicating no overall lack of fit for the model.

Another comparison of the two models can be made in terms of the estimates of the sensitive trait proportions and their respective standard errors at each of the five covariate levels. The results are reported in Table 1. In addition the table reports the SE ratio, i.e., the ratio of the standard error of the weighted average of the sensitive trait proportion for the nonmonotonic model to the same

standard error for the standard covariate model.

Table 1. Estimates of Sensitive Trait Proportions for Contrived Data

Cov Level	n	Model	Estimate	SE
1	13	Monotonic	.338	.196
		Non-Monotonic	.112	.046
2	48	Monotonic	.338	.124
		Non-Monotonic	.278	.046
3	78	Monotonic	.338	.088
		Non-Monotonic	.512	.082
4	48	Monotonic	.338	.124
		Non-Monotonic	.256	.046
5	13	Monotonic	.338	.196
		Non-Monotonic	.101	.045
Weighted Avg.		Monotonic	.338	.057
		Non-Monotonic	.342	.036
(μ (merge point) = 2.95 for the nonmonotonic model)				
SE Ratio			.63	

The results for the contrived data can be summarized as: (a) the nonmonotonic model shows no misspecification while the monotonic model shows clear misspecification, (b) the nonmonotonic model shows no lack of fit to the data and (c) the nonmonotonic model shows consistently lower standard errors than the monotonic model. The data in this example was contrived to be nonmonotonic and thus it is no surprise that the "merged" logistic gives superior results. In the next section we investigate an actual set of data from a RRT survey using the covariate technique.

5. USE OF MODEL WITH ACTUAL DATA

Four different covariates were used in a Warner model to estimate the tendency for business executives to bluff when announcing a price increase. The items were obtained from a survey of business executives done at a major Eastern university. The four covariates investigated were: (a) the strength of the company's product in its industry, (b) the strength of an executive's information about the market conditions of his/her industry, (c) the extent to which the signal (that a price increase was forthcoming) could be reversed by the executive, (d) the extent to which firms face entry and exit barriers within their respective industries. Each of these covariates were obtained through a survey of firm characteristics completed by the executives. Approximately 550 executives participated from a broad range of industries. The following pair of questions were asked of each executive:

1. Think of the most recent announcement made by your firm (or a unit of it) concerning a price increase.

To my knowledge we intended to do what we announced: it was not a bluff.

___ yes ___ no

2. Think of the most recent announcement made by your firm (or a unit of it) concerning a price increase.

To my knowledge the announcement was a bluff.

___ yes ___ no

The probability of selecting the sensitive item, i.e, item #2, was fixed at .30 by requesting the respondent to withdraw a dollar bill from his/her wallet and check the last digit of its serial number. If the last digit of the serial number was 8,9, or 0, the executive was instructed to answer item #2; otherwise, he/she was instructed to answer item #1. The covariate scores were obtained through an attached survey. Five levels for each covariate were obtained by dividing the data into approximate quintiles.

The data from the survey was subjected to analyses by both (1) the monotonic covariate model and (2) the nonmonotonic ("merged" logistic) model. Table 2 presents the parameter estimates for each of the models and their standard errors. Table 3 provides the sensitive trait proportions and their associated standard errors for each model at each level of all four covariates. Additionally, the weighted averages, their standard errors and the SE ratio is presented.

Table 2. Parameter Estimates for Monotonic and Nonmonotonic Models

Covariate	Monotonic Model		Nonmonotonic Model	
	α	β	α	β
1. Product Strength	-6.30 (3.94)	1.1 (.845)	2.4 (2.15)	-7.15 (4.43)
2. Info Strength	-11.20 (20.2)	1.9 (4.07)	-4.9 (2.04)	1.57 (1.05)
3. Reversibility	- 1.07 (1.60)	-.65 (.917)	-11.8 (5.09)	5.07 (2.41)
4. Barriers	-9.14 (9.28)	1.57 (1.91)	-5.41 (1.78)	1.84 (.846)

Table 3. Estimates of Sensitive Trait Proportions (Bluffing)

Covariate Level	Model	Covariate: 1. Product Strength $\mu = 4.55$	2. Info Strength $\mu = 3.00$	3. Reversibility $\mu = 3.15$	4. Barriers $\mu = 2.75$
1	Monotonic	.005 (.017)	.000 (.002)	.152 (.116)	.001 (.004)
	Nonmonotonic	.000 (.000)	.148 (.037)	.299 (.055)	.102 (.035)
2	Monotonic	.016 (.036)	.001 (.007)	.086 (.068)	.002 (.014)
	Nonmonotonic	.000 (.000)	.035 (.034)	.003 (.006)	.018 (.019)

3	Monotonic	.047 (.066)	.004 (.033)	.047 (.069)	.012 (.042)
	Nonmonotonic	.000 (.001)	.007 (.015)	.000 (.000)	.007 (.011)
4	Monotonic	.130 (.083)	.027 (.105)	.025 (.058)	.055 (.092)
	Nonmonotonic	.173 (.052)	.035 (.034)	.001 (.002)	.043 (.031)
5	Monotonic	.310 (.130)	.158 (.120)	.025 (.058)	.055 (.092)
	Nonmonotonic	.300 (.053)	.148 (.037)	.085 (.053)	.221 (.051)
Wgt. Avg.	Monotonic	.092 (.032)	.037 (.033)	.063 (.034)	.050 (.032)
	Nonmonotonic	.085 (.014)	.070 (.014)	.078 (.015)	.068 (.013)
SE Ratio		.44	.42	.44	.41

Note: SE's are given parenthetically; SE ratio is the ratio of the wgt avg standard error for the nonmonotonic model to that of the monotonic model.

From the above results, it is clear that the nonmonotonic model consistently yields lower standard errors than the monotonic model. The SE ratio never exceeds .44 for any of the covariates. In addition, the nonmonotonic model yields significant estimates for beta in two of the four models, i.e, the "Barriers" covariate (z= 2.18) and the "Reversibility" covariate (z = 2.10), while the monotonic model is misspecified for all four covariates. In two of the four cases, i.e, the "Info Strength" and "Product Strength" covariates, the nonmonotonic model is also seemingly misspecified. The sensitive trait may have no relationship to the covariate in these cases, or the relationship may be fundamentally different from that of a merged logistic.

6. CONCLUSIONS

The user of covariate RRT must be alert to the possibility of nonmonotonic relationships between the covariate and the sensitive trait proportion. If evidence exists for the misspecification of the standard (monotonic) covariate model, the RRT researcher should investigate the possibility of other covariate models. While this study has not developed a general nonmonotonic covariate RRT model, it has shown the potential utility of alternatives to the standard covariate approach. In the future, RRT users may find the use of other functional types, e.g., a beta function, to be helpful in reducing the standard errors of sensitive trait estimates. Additionally, composite functional types may be usefully employed via "spline" functions.

This study has shown the reasonableness of attempting a nonmonotonic model through the demonstrated capacity of the "merged" logistic function to estimate the sensitive trait proportions more precisely. A first step in any covariate RRT application should be the fitting of separate noncovariate, e.g., Warner or unrelated question models, at each of the covariate levels. A graph of the estimated sensitive trait proportions at each covariate level may well suggest the untenability of the standard (monotonic) covariate treatment of the problem.

REFERENCES

Greenberg, B. G., Abul-Ela, A., Simmons, W. R., and Horvitz, D. G., (1969), "The Unrelated Question Randomized Response Model: Theoretical Framework." *Journal of the American Statistical Association*, 64, 520-539.

Rao, C.R. (1965) *Linear Statistical Inference and its Applications.* New York: John Wiley

Scheers, N.J. and Dayton, C.M. (1988) "Covariate Randomized Response Models." *Journal of the American Statistical Association*, 83, 969-974.

Sudman, S. and Bradburn, N. M. (1974), *Response Effects in Surveys*, Chicago: Aldine.

Warner, S. L. (1965), "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association*, 60, 63-69.